IBM® Lotus® Sametime® Reference Architecture in a VMware® Infrastructure 3 Environment
Real-time collaboration has become a vital component of on-demand business, and IBM Lotus Sametime has quickly become an important component in many IBM Lotus Notes® and IBM Lotus Domino® environments. Typical Lotus Sametime deployments that support several thousand users require multiple servers in the physical environment, which results in high capital and operating costs. Virtualization offers unprecedented opportunities in commodity server consolidation, workload management, and return on investment.

VMware Infrastructure 3 does the following:

- Consolidates the server sprawl in Sametime physical environments into fewer servers.
- Provides a dynamically adapting, highly available solution to deploy IBM Lotus Sametime that is not possible in physical environments.

Together, IBM and VMware deployed and proved that the complex configuration of IBM Lotus Sametime 7.5.1 can scale on VMware ESX Server 3.0.2, using one powerful machine. This white paper validates this solution and provides the reference architecture to deploy it in production. The test results clearly demonstrate that a Lotus Sametime deployment running in a VMware Infrastructure 3 environment can successfully support 100,000 Lotus Sametime chat users while maintaining the service level agreement (SLA) requirement for end-user response time.

The following topics are discussed in this paper:

- “Introduction” on page 2
- “Objectives and Methodology” on page 3
- “Test Configuration” on page 4
- “Test Results” on page 6
- “More Observations” on page 10
- “Conclusion” on page 10
- “References” on page 10

### Introduction

IBM and VMware continue to collaborate on leveraging VMware Infrastructure 3 to offer a solid, robust virtualization platform to deploy IBM's software products on. Due to the flexibility of virtual platforms, running IBM Lotus Sametime in a VMware virtual machine can provide dramatic price and performance ratios. These platforms are designed to enhance server consolidation and disaster recovery.
IBM Lotus Sametime is the market-leading platform for unified communication and collaboration and offers secure integrated enterprise instant messaging, VoIP, video chats, and Web conferencing capabilities with the security features required for business use. The Lotus Sametime server also supports open standard interfaces, such as the Gateway component to support public instant messaging providers, and is integrated with standard telephony components.

VMware Infrastructure is the most widely deployed, production-ready software suite available on commodity hardware to optimize and manage IT environments through virtualization. Using virtualization-based distributed infrastructure services, such as VMware VMotion™, VMware High Availability, and VMware Distributed Resource Scheduler, VMware Infrastructure 3 provides transformative cost savings to the dynamic data center while increasing serviceability, efficiency, and reliability.

In this test, we deployed a complex IBM Lotus Sametime 7.5.1 configuration, using ESX Server 3.0.2, to verify whether this solution can support several thousand users on a single powerful machine. On this system, we ran the IBM Lotus Instant Messaging workload. This benchmark simulates Instant Messaging activities, such as logging in, chatting, adding other users to contact lists, logging out, and so on.

We ran the test for 20 hours with the goal of maintaining the end-user response time as stated in the SLA. Customers who consider instant messaging mission critical will accept up to a two-second delay in message delivery. By running over 100,000 Lotus Sametime chat users simultaneously, we demonstrated that IBM Lotus Sametime on ESX Server offers a production-ready environment while effectively using the underlying physical server and storage resources. This white paper describes the results obtained from these tests, which can help systems engineers and IT architects determine an effective workload throughput and virtual machine usage for their data centers.

**Objectives and Methodology**

The objective of this test was to verify whether a complex configuration of IBM Lotus Sametime 7.5.1 software can scale on ESX Server 3.0.2, using a powerful machine. We targeted 100,000 concurrent chat users on a single ESX Server.

The testing simulated a real-world environment where multiple virtual machines, comprising the different components of the infrastructure, are running on one ESX Server. This unique solution is not possible in the physical world. To support a similar configuration in a physical environment, three to five physical servers would be required to scale beyond 35,000 concurrent users. In a production environment, this configuration should be hosted on multiple ESX Servers for fault tolerance and disaster recovery.

We used the IBM Lotus Sametime chat or Instant Messaging workload for the tests, but we did not test the instant meeting (collaboration) or scheduled meeting (Web conference) workloads.

**Methodology**

The instant messaging workload simulates the following activities:

- An active, instant messaging user logging in and chatting with another user in rich-text mode
- Exchanging about five messages
  - Each message was approximately 512 bytes long with a 30-second gap between each message.
- The workload adding a user to the Contacts list every 24 iterations
- Logged out, waiting about ten minutes, and logging in every 12 iterations

The iteration values are configurable, and the specified values help illustrate the example. IBM Lotus Notes 7.5 utility Server.Load provides the ST75IM script, which performs an activity that is similar to what was performed for this benchmark.

We modified the sametime.ini file to increase the capacity of the Sametime Multiplexers to allow more than 20,000 connections. Multiple drivers, approximately ten to twelve, were used to simulate over 100,000 Sametime users during these tests.
We gradually increased the number of simulated Sametime users, while keeping the average response time under two seconds. If user response time is greater than two seconds, the results were considered unacceptable.

We used esxtop, a performance-monitoring tool for ESX Server, and VMware VirtualCenter performance data to record ESX Server and virtual machine-related performance counters.

Test Configuration

Hardware configuration

We used the IBM System X™ 3650 server in the test:

Table 1 provides the hardware configuration in the test environment.

<table>
<thead>
<tr>
<th>Component</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>2x quad-core 2.33GHz Intel “Clovertown” processors</td>
</tr>
<tr>
<td>Memory</td>
<td>24GB</td>
</tr>
<tr>
<td>Internal Disks</td>
<td>2x 73GB SCSI</td>
</tr>
<tr>
<td>Ethernet Adaptors</td>
<td>2x 10/100/1,000 Mbps internal adapters</td>
</tr>
<tr>
<td>Fibre Channel HBA</td>
<td>QLogic QLA2432 host bus adaptor (HBA)</td>
</tr>
</tbody>
</table>

The QLogic adaptors provided connectivity to the IBM DS4800 storage area network (SAN).

The client system that ran the instant messaging workload needs 2GB RAM and 1-2 CPUs, depending on the number of simulated users. We used between 10 to 12 driver machines, and each driver handled between 8,000 to 10,000 users. Each driver machine used two 2.4GHz or higher Intel® Xeon® processors and 2GB of RAM. The driver machines ran Microsoft® Windows Server® 2003 Enterprise Edition (32-bit) and the IBM Lotus instant messaging workload.

Storage Configuration

In the test system, we did the following:

- Connected one IBM DS4800 storage SAN with two controllers, each with 8GB of cache using Fibre Channel Protocol.
- Attached over 15 36GB 15K RPM disks to the storage server
  
  All virtual disks were on the same VMFS LUN.

Table 2 provides the storage configuration in the test environment.

<table>
<thead>
<tr>
<th>Component</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN</td>
<td>IBM DS4800 SAN</td>
</tr>
<tr>
<td>Disks</td>
<td>15 x 36GB 15,000 rpm</td>
</tr>
<tr>
<td>LUNs</td>
<td>500GB RAID-5 Each virtual machine 20GB</td>
</tr>
</tbody>
</table>

Software Configuration

In the test system, we did the following:

- Installed ESX Server 3.0.2 on one of the internal hard disks.
Created four virtual machines using VirtualCenter 2.01. Each virtual machine used Microsoft Windows 2003 Enterprise Edition SP1 (32-bit) as the guest operating system.

Configured two virtual machines as IBM Lotus Sametime Community servers and the other two as Lotus Sametime Multiplexers.

Created an IBM Lotus Sametime cluster using the two Lotus Sametime Community servers.

To create an IBM Lotus Sametime cluster, use an IP sprayer or equivalent solution. We used IBM WebSphere® Edge Network Dispatcher. Another option is to use a DNS round-robin algorithm to route the user requests. This is not a robust option for production deployments but does offer a less expensive alternative for testing.

Used a clustered IBM Lotus Domino LDAP 7.0.1 directory configured on a remote server.

No pictures were present in this directory. If pictures are present because more data needs to be transferred, the load on the IBM Lotus Sametime server will increase.

We used default values for all configuration parameters. Table 3 provides the software configuration in the test environment.

<table>
<thead>
<tr>
<th>Software</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESX Server</td>
<td>3.0.2</td>
</tr>
<tr>
<td>VirtualCenter</td>
<td>2.01</td>
</tr>
<tr>
<td>Microsoft Windows</td>
<td>Microsoft® Windows® 32 bit</td>
</tr>
<tr>
<td></td>
<td>2003 Standard Edition SP1</td>
</tr>
<tr>
<td>IBM Lotus Domino</td>
<td>7.0.2 FP2</td>
</tr>
<tr>
<td>IBM Lotus Sametime</td>
<td>7.5.1</td>
</tr>
<tr>
<td>IBM Lotus Domino LDAP</td>
<td>7.0.1</td>
</tr>
</tbody>
</table>

Figure 1 represents the test environment.
Table 4 provides the virtual machine configurations in the test environment.

<table>
<thead>
<tr>
<th></th>
<th>Lotus Sametime Community Server</th>
<th>Lotus Sametime multiplexers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual CPU</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Virtual Memory</td>
<td>4GB</td>
<td>2GB</td>
</tr>
<tr>
<td>IBM Lotus Software</td>
<td>IBM Lotus Domino 7.0.2 FP2</td>
<td>IBM Lotus Sametime 7.5.1</td>
</tr>
<tr>
<td>Virtual Disk</td>
<td>20GB</td>
<td>20GB</td>
</tr>
</tbody>
</table>

The latest version of VMware Tools was installed in each guest operating system.

**Test Results**

We provide detailed test results for the 100,000 Sametime users solution deployed in the VMware Infrastructure 3 environment. The results clearly validate that a complex configuration of IBM Lotus Sametime 7.5.1 software can scale on ESX Server 3.0.2, using one powerful server, while providing excellent user response time.
Figure 2 shows the average response time experienced by the 100,000 users during the 20-hour test run. As the figure illustrates, 99.22 percent of users reported an average response time of less than one second.

**Figure 2. Average End User Response Time for the Sametime users**

![Response Time Distribution](image)

Figure 3 and Figure 4 provide additional details regarding the peak number of transactions performed on one of the Community Servers during the 20-hour test run. A similar profile was also observed on the other Community Server. IBM Lotus Sametime servers must be sized for peak situations because the server load during this time is usually greater than the load during the steady state. Although this is true for most server workloads, this is especially important for Sametime servers, which usually run at low usage levels during the steady state. Peak load for IBM Lotus Sametime Connect clients is the period in which the maximum number of Sametime clients are connecting to the server. Figure 3 illustrates that, during the test, one community server in a virtual environment could support at least the following for each 5-minute interval during the test:

- 57,000 concurrent logins
- 320,000 instant messages
- 17,000 instant messaging channels
- 525,000 status changes
Figure 4 shows that one community server in a virtual environment supported at least the following for each 5-minute interval during the test:

- 500 concurrent network chats
- Creation of over 1600 Buddy List Channels
- Over 1600 successful logins

Figure 5 and Figure 6 illustrate the corresponding peak CPU usage for the four virtual machines. The peak CPU usage for the first IBM Lotus Sametime Community Server averaged about 20% with a maximum value of 57%. For the second Sametime Community Server, the peak CPU usage averaged about 17% with a maximum value of 39%.
The peak CPU usage for the first Sametime Multiplexer averaged about 15% with a maximum value of 25%. For the second Sametime multiplexer, the peak CPU usage averaged about 14% with a maximum value of 24%.

The following list provides additional performance data for the test. Each data point provides average operations per minute that occurred during the Benchmark:

- 419 users logged in
- 900 users change status, resulting in 90,000 status change notifications
- 3,160 IM channels created
IBM® Lotus® Sametime® Reference Architecture in a VMware® Infrastructure 3 Environment

- 4,714 small Instant Messages sent
- 3,160 large Instant Messages (40 KB; i.e., screen capture) sent
- 900 User Info requests (business card)
- 400 reads and 267 updates made from/to storage (buddylist 30 KB)
- 16 n-way (3 or more parties) chats are started

More Observations

In this benchmark, we used a clustered remote LDAP directory. When using local Lotus Domino Directory, CPU usage could be higher because authentication is performed locally. Lotus Sametime instant messaging servers typically scale well and are not heavy on network usage. Meetings are more network and CPU intensive.

Using separate multiplexers can dramatically increase the scalability of the environment. Multiplexers do not need the same amount of memory and virtual CPUs as the Lotus Sametime Community Servers.

Conclusion

This white paper supports the conclusion that IBM Lotus Sametime is an excellent candidate for virtualization. Running IBM Lotus Sametime server in virtual machines provides an effective production-ready instant messaging platform for the enterprise. This document validates the scalability of a complex configuration of IBM Lotus Sametime 7.5.1 on ESX Server 3.0.2, using a powerful server, and provides the reference architecture to deploy such a solution. The test results clearly demonstrate that Lotus Sametime deployment in a VMware Infrastructure 3 environment can, for an extended period of time, successfully support 100,000 Lotus Sametime users with less than a two-second response time.

References

IBM

- Sametime 7.5.1: Best Practices for Enterprise Scale Deployment
  http://www.redbooks.ibm.com/abstracts/sg247410.html
- Lotus Domino products are officially supported on ESX Server
  http://www-1.ibm.com/support/docview.wss?rs=203&uid=swg21106182
- How to size Domino and Sametime systems for full production loads on ESX Server
  http://www-1.ibm.com/support/docview.wss?rs=203&uid=swg21267837
- Idle Sametime server on Linux: CPU usage on ESX host server

VMware

- Performance Tuning Best Practices for ESX Server 3